CLAIMS

What is claimed is:

1. A method, for use in query optimization in a relational database management system, said method comprising the steps of:

(a) generating statistical information regarding data which represents the results of an operation involving one or more columns of a database;

(b) deriving a statistical soft constraint from said statistical information that reflects a statistical property of said data; and

(c) using said statistical soft constraint to estimate a cardinality value for the result of applying one or more query predicates in a query plan.

2. The method of claim 1 further comprising the step, prior to step (a), of creating a materialized column containing said data, wherein said data comprises the results of said operation involving one or more columns of a database.

3. The method of claim 2 wherein said materialized column is stored in the database.

4. The method of claim 2 wherein said statistical soft constraint comprises a constraint predicate and an associated probability value, said associated probability value reflecting the percentage of rows of said one or more columns for which said constraint predicate is true.

5. The method of claim 2 wherein said step of generating statistical information comprises gathering said statistical information regarding said data utilizing a statistics gathering process provided by the relational database management system.

6. The method of claim 1 wherein the step of generating statistical information comprises analyzing the data using an SQL statement.

7. The method of claim 6 wherein said SQL statement groups a selection to obtain frequencies.

8. The method of claim 1 further comprising the step, prior to step (b), of analyzing said statistical information and determining a useful subset of said statistical information from which to derive said statistical soft constraint.

9. The method of claim 1 wherein said statistical soft constraint comprises a constraint predicate and an associated probability value, said associated probability value reflecting the percentage of rows of said one or more columns for which said constraint predicate is true.

10. The method of claim 9 wherein said query predicate comprises an expression involving two different columns.

11. The method of claim 10 wherein the step (c) of using said statistical soft constraint comprises the steps of:

(c1) normalizing said query predicate, if necessary, such that the right-hand side of said query predicate expression comprises a constant;

(c2) determining whether said query predicate matches said constraint predicate;

(c3) setting a selectivity for said query predicate equal to said associated probability value if said query predicate matches said constraint predicate; and

(c4) setting a selectivity boundary for said query predicate based upon said associated probability value if said query predicate does not match said constraint predicate.

12. The method of claim 9 wherein said query predicate comprises an operation upon a column.

13. The method of claim 12 wherein the step (c) of using said statistical soft constraint comprises the steps of:

(c1) normalizing said query predicate, if necessary, such that the right-hand side of said query predicate expression comprises a constant;

(c2) determining whether said query predicate matches said constraint predicate;

(c3) setting a selectivity for said query predicate equal to said associated probability value if said query predicate matches said constraint predicate; and

(c4) setting a selectivity boundary for said query predicate based upon said associated probability value if said query predicate does not match said constraint predicate.

14. The method of claim 9 wherein said query predicate comprises two predicates, the first predicate involving a first column and the second predicate involving a second column, said first column being a different column from said second column, wherein said constraint predicate comprises and expression including said first column and said second column.

15. The method of claim 13 wherein the step (c1) of using said statistical soft constraint comprises the steps of:

(c1) normalizing said constraint predicate, if necessary, to produce a normalized constraint predicate wherein the left-hand side of said normalized constraint predicate comprises said first column;

(c2) substituting occurrences of said first column in said first predicate with the right-hand side of said normalized constraint predicate, such that said first predicate only refers to said second column;

(c3) transposing said first predicate, if necessary, to produce a transposed first predicate wherein the left-hand side of said transposed first predicate comprises said second column; and

(c4) setting a selectivity or selectivity bound based upon said transposed first predicate, said second predicate and statistical information regarding said second column.

16. A database management system comprising:

means for generating statistical information regarding data which represents the results of an operation involving one or more columns of a database;

means for generating a statistical soft constraint using said statistical information; and

means for utilizing said statistical soft constraint to estimate a cardinality value for the result of applying one or more query predicates in a query plan.

17. The database management system of claim 16 wherein said statistical soft constraint comprises a constraint predicate and an associated probability value reflecting the percentage of rows of said one or more columns for which said constraint predicate is true.

18. The database management system of claim 17 wherein said means for utilizing comprises:

means for identifying a type of said query predicate;

means for normalizing said query predicate and said constraint predicate;

means for comparing said query predicate with said constraint predicate;

means for setting a selectivity equal to said probability value when said query predicate matches said constraint predicate; and

means for setting a selectivity bound based upon said probability value when said query predicate does not match said constraint predicate.

19. The database management system of claim 16 wherein said query predicate comprises a first and second predicate.

20. The database management system of claim 19 wherein said means for utilizing further comprises:

means for generating a twin predicate from said first predicate; and

means for setting a selectivity or selectivity bound based upon said twin predicate, said second predicate and said probability value.

21. A computer program product comprising:

(a) a computer readable medium;

(b) code means contained in said medium for instructing a computer to perform the steps of:

(i) generating statistical information regarding data which represents the results of an operation involving one or more columns of a database;

(ii) deriving a statistical soft constraint from said statistical information that reflects a statistical property of said data; and

(iii) using said statistical soft constraint to estimate a cardinality value for the result of a query predicate in a query plan.

22. The computer program product of claim 21 wherein said computer readable medium

is chosen from the group consisting of a modulated electrical signal, a modulated optical signal, a magnetic storage medium and an optical storage medium.

23. A computer readable medium containing program instructions for use in query optimization in a relational database management system, said program instructions for:

(a) generating statistical information regarding data which represents the results of an operation involving one or more columns of a database;

(b) deriving a statistical soft constraint from said statistical information that reflects a statistical property of said data; and

(c) using said statistical soft constraint to estimate a cardinality value for the result of applying one or more query predicates in a query plan.